

Citation for published version:

Finnegan, D, O'Neill, E & Proulx, M 2017, An Approach to Reducing Distance Compression in Audiovisual Virtual Environments. in *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)* ., 7901607, IEEE. <https://doi.org/10.1109/SIVE.2017.7901607>

DOI:

[10.1109/SIVE.2017.7901607](https://doi.org/10.1109/SIVE.2017.7901607)

Publication date:

2017

[Link to publication](#)

Publisher Rights

Unspecified

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

An Approach to Reducing Distance Compression in Audiovisual Virtual Environments

Daniel J. Finnegan*

Eamonn O'Neill†

Michael J. Proulx‡

University of Bath

ABSTRACT

Perception of distances in virtual reality (VR) is compressed: objects are consistently perceived as closer than intended. Although this phenomenon has been well documented, it is still not fully understood or defined with respect to the factors influencing such compression. This is a problem in scenarios where veridical perception of distance and scale is essential. We report the results of an experiment investigating an approach to reducing distance compression in audiovisual VR based on a predictive model of distance perception. Our test environment involved photorealistic 3D images captured through stereo photography, with corresponding spatial audio rendered binaurally over headphones. In a perceptual matching task, participants positioned an auditory stimulus with respect to the corresponding visual stimulus. We found a high correlation between the distance perception predicted by our model and how participants perceived the distance. Through automated manipulation of the audio and visual displays based on the model, our approach can be used to reposition auditory and visual components of a scene to reduce distance compression. The approach is adaptable to different environments and agnostic of scene content, and can be calibrated to individual observers.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality; H.5.m [HCI]: Miscellaneous

1 INTRODUCTION

Egocentric distance perception is defined as the perception of distance between oneself and an object [23]. In VR the subjective perception of distance is frequently compressed (i.e. perceived closer) compared to the real world [12, 16, 29]. In entertainment applications veridical perception may not be critical, although it has been shown to be an important factor in preference and enjoyment [17]. However, applications such as virtual museum tours, actions involving reaching to nearby objects, and augmented remote medical surgery (telerobotic surgery) can benefit from a perception of space as close as possible to the real world [13, 19, 22].

Distance compression is both a unimodal and a crossmodal phenomenon, and has been studied extensively in audio, visual and audiovisual environments. For example, Renner et al. segmented the space around an observer into three cue regions: personal, action and vista space [23]. Each space is defined by the distance from the observer, and different cues (e.g. binocular disparity, occlusion, relative size) are applicable in different spaces. Some cues overlap and apply in more than one cue region. With respect to audition, Zahorik studied distance cues in virtual environments [30]. His assessment of various cues such as sound intensity, binaural differences, direct-to-reverberant ratio, and spectral content demonstrated

a similar context-sensitive appropriation of distance cues to that in Renner et al.

A well known example of crossmodal influence on perception is the ‘Ventriloquist Effect’, where the source location of speech auditory stimuli is often localised to the nearest face in the visible environment [1, 6, 18]. Ernst & Banks demonstrated that, for combinations of visual and haptic stimuli, humans perceive in a ‘statistically optimal fashion’, termed Maximum Likelihood Theory [8]. Maximum Likelihood theory has been demonstrated with vision and audition [3] but the authors note that their results do not rule out ‘visual capture’, a theory suggesting an overall bias towards the visual modality. In order to come to a consensus, the study reported here investigates crossmodal influence on distance perception in audiovisual virtual environments using a perceptual matching task.

2 BACKGROUND

With respect to audition, the energy in reverberant environments lingers longer than in anechoic environments due to the reflections and echoes. The human auditory system is understood to integrate the intensity level of a sound source, with the energy in its reflections, in order to estimate the distance to the sound source [31]. Zahorik conducted an analysis of how humans weigh intensity and direct-to-reverberant energy ratio cues (D-R) when determining distance [30]. He found that the weight attributed to each cue when making distance judgements changes as a function of sound source type as well as sound source position. When sounds were presented directly in front of participants, intensity was weighted more than D-R, meaning participants relied on the intensity of the direct wave rather than the reverberant energy in the environment to make judgements. However, as sounds were moved further to the lateral side of the listener, participants began to weigh intensity less and rely more on D-R. Therefore, the efficacy of auditory cues correlates with their angular displacement of the sound source relative to the observer. The changes in weighting between the cues results in repercussions for interventions that try to exploit particular cues in order to relieve distance compression. This could lead to localization errors in depth: any compression compensation method would need to accommodate this. It is also unclear how these weights may interact with visual perception in an audiovisual environment.

Distance perception can also result from internal processing. Distance information is extracted from cues and integrated, exploiting the cue context. For example, speech is a well known signal and has certain expected attributes. A speech source is expected to originate from the mouth of a person, and is thus expected to travel outwards, in multiple directions (due to the propagation of sound in free space), however, it is expected to propagate in a particular direction most strongly, i.e. towards the listener who is the target of the speech. This expectation of the sound source origin has repercussions for vision; the listener would expect to see an appropriate visual cue emanating from the same source. However, the Ventriloquist Effect may be exploitable. Decomposing an audiovisual stimuli into auditory and visual components, and repositioning them may still result in the stimuli being perceived as a single object. All temporal properties of the stimuli remain unaffected, simply their *spatial* alignment is changed. This idea has been previously applied by Finnegan et al. in

*e-mail: djf32@bath.ac.uk

†e-mail: eamonn@cs.bath.ac.uk

‡e-mail: m.j.proulx@bath.ac.uk

a simple abstract environment with minimal distance cues [9]. We test whether the effect holds true in natural, cue rich environments.

Crossmodal binding has been studied to observe visual or audio bias effects on distance perception [2, 5, 14]. Chan et al. took distance judgements in an audiovisual target selection task. Visual stimuli consisted of a picture of a face being lit up, while auditory stimuli consisted of a speech signal of the phrase ‘Hi’. All stimuli emanated from a 2D circular array consisting of a foreground arc and a background arc. The front arc was situated 60 cm from the listener and the background arc was 120cm. They measured localisation accuracy, asking participants to choose whether an audio, visual, or audiovisual stimulus appeared in either the background or foreground arc while staring at a fixation point at the 0° position. They found an interaction effect between the audiovisual modality and the presentation angle along the arc. Accuracy across audiovisual *incongruent* trials was greater in the central location than the periphery. Audiovisual *congruent* trials were as accurate in the central locations as the peripheral [5]. This suggests that the auditory stimulus was a distraction and did not result in more accuracy than the visual only condition. However, participants were told to ignore the auditory stimulus in the audiovisual conditions. Results could have been different if participants were told to perceive the audiovisual stimuli as a single multimodal stimulus, as multimodal processing is broadly considered to be more effective than unimodal processing (See [25] for example).

Anderson & Zahorik took distance judgements with auditory, visual, and audiovisual photoreal stimuli in the frontal plane [2]. The expected compression of distance occurred in all 3 modalities, and in agreement with Chan et al. they found AV and V conditions were not significantly different from one another, yet visual cues resulted in less variability across trials. They concluded that the presence of visual information improves accuracy in distance perception. Their design consisted of a captured binaural room impulse response (BRIR) for naturalistic spatial audio, and a photo of the loudspeaker as auditory and visual stimuli respectively. Similarly to the study by Chan et al. participants would have made the implicit association between the crossmodal stimuli and bound the two together.

Considering that this crossmodal binding result is stronger for spatially congruent stimuli, an interesting question occurs: ‘What are the bounds of congruency within which perceptual binding can occur with respect to distance perception?’. Gorzel et al. investigated the bounds of incongruence for a range of audiovisual stimuli [10]. In a perceptual study involving stereoscopic imagery viewed using polarised filter based 3D glasses and a large HDTV monitor, with ambisonic audio rendered over headphones, they found that there were margins of misalignment for which audiovisual stimuli were perceived as consistent. For visual stimuli positioned 2m from the observer, 50% (chance level) estimates were consistently made with a margin of 1m. However, when the visual stimuli was pushed back to 8m, a margin of 3m resulted in 50% performance. These trends suggest that the margin of misalignment increases logarithmically with the egocentric distance of the visual stimulus. This implies that further distance allows for greater incongruency while ensuring perceptual binding.

3 SOLVING THE DISTANCE COMPRESSION PROBLEM

In the visual domain, manipulating the geometry of the scene can reduce the level of distance compression in blind walking tasks [15]. The technique, known as minification, involves shrinking the image to be displayed, and then rendering the resulting image so that the complete field of view is scaled appropriately. Minification adjusts the angle of declination (see [21]) with respect to a scaling factor as in Equation 1.

$$\theta_{new} = \arctan(m \cdot \tan \theta_{orig}) \quad (1)$$

Kuhl et al. conducted a between-subjects experiment investigating the effect of ‘minifying’ the scene geometry on distance percep-

tion, and found that compression was reduced with the minification rendering technique compared to a control condition with an unminified scene [15]. However, compression still occurred regardless of minification.

Steinicke et al. experimented with redirected walking, a technique to lead a user immersed in a virtual world along a specific trajectory [26]. This enables virtual world designers to expand the exploration space in the virtual world to a greater area than is physically available. The most striking finding is that users can be made to walk in a circle of radius greater than 22 meters yet feel as if they have maintained a straight line. An added benefit of redirected walking is that high cognitive load has an impact on the curve at which participants could be redirected [4]. In hectic VR environments, observers may be very easily led to believe they are immersed in a bigger environment than physical constraints allow.

In implementing their redirected walking technique, Steinicke et al. found that distances can also be modified in the virtual environment so that they don’t match the physical world, yet are perceptually veridical to the user [26]. Finnegan et al. found a similar effect with incongruent audiovisual environments, where the variance of the point of subjective equality (PSE) was as great as 0.5 meters when audio and visual stimuli were misaligned [9]. Their incongruence function determines the position of an audio source given a visual anchor as in Equation 2.

$$\phi = \left(\frac{\hat{y}}{k} \right)^{\frac{1}{\alpha}} \quad (2)$$

By intentionally misaligning the acoustic source from a visual stimulus (i.e. seeing and hearing a loudspeaker), participants were more accurate in their estimates of distance based in a 2 alternative forced choice (2AFC) task¹.

4 HYPOTHESIS

The aim of this current study was to assess incongruence as a viable solution to the distance compression problem in photorealistic virtual environments. In summary, we note the following points made in this background review:

1. Distance perception within virtual environments has been found to be compressed with respect to the real world across various studies.
2. Distance perception is seen as a crossmodal cognitive task and therefore should be studied from the perspective of multi-sensory integration and psychophysics.
3. It is unclear whether incongruency is only effective in *some* virtual environments, consisting of abstract and unfamiliar stimuli, or whether the positive impact of incongruence is generalizable and applicable to more natural environments.

The first and second points have been discussed in the preceding sections. In order to address the third point, we designed and conducted an experiment which looked at distance perception in virtual audiovisual environments. We anticipated that incongruence would be effective in environments that provide more distance cues to the observer; in this instance, extra cues in the form of relative size, occlusion, binocular disparity and shading.

H1: Incongruence would be effective in predicting distance perception in environments that provide more distance cues to the observer.

¹ See [11] for details on AFC tasks.



Figure 1: The camera rig used to take the stereo photographs.

5 EXPERIMENT DESIGN

The experiment involved 40 participants (12 Female), a mixture of undergraduate and postgraduate students at a local university, with a mean age of 26 years. The task was to match the position of a car horn to the spatial position of a target car. Spatial audio was implemented using the same custom plugin written for the Unity Game Engine used by Finnegan et al. [9]. The plugin is publicly available for download: [<https://github.com/Ps2Fino/Unity-Audio-Plugin>]. All participants were unaware of the experiment’s hypothesis, and had no known hearing impairments. The Oculus Rift headset was calibrated based on the IPD of the individual participant using the configuration utility from the Oculus Rift SDK. Participants were seated in a chair, with head movement restricted using a chin rest.

5.1 Environment and Stimuli

Inside the virtual environment, participants saw a series of scenes each depicting a row of cars parked, viewed from the side. In each scene, the closest car was recorded as being 5 meters from the camera’s position, measured using a tape measure. Each scene was rendered from stereo photographs captured with a Canon EOS 60D, with a 50mm lens, mounted on a custom camera rig consisting of a tripod, sliding dolly, and wooden frame. The photographs were captured by taking a single photograph, namely the photograph for the right eye, then shifting the camera 2.5 inches to the left and taking a second photograph. 2.5 inches was chosen as it is a measured estimate for the IPD across the general population [7]. A sample image pair from our asset set is shown in Figure 2. Scenes 1 & 3 contained 6 cars while scene 2 contained 5 cars for a total of 17 distances across all scenes. The audio stimulus was a synthesized car horn (1 second, mono channel) that participants moved around the scene. All scenes were comparable in terms of visual cues, and the horn sound was identical in all trials.

5.2 Procedure

Before beginning the experiment, each participant was guided through a tutorial that demonstrated the input mechanism and the task for the experiment. The UI shown in Figure 3 was displayed throughout the entire experiment to remind participants. After the tutorial, the main experiment began immediately. Using the game controller, participants could move the current location of the car horn in a semi-circle arc in front of them, from 0° - 180° . No online auditory feedback was provided; at any time, participants could poll the audio in order to hear the horn from its current position. They could also hear the horn from a reference position 1 meter directly in front of them. When they felt the position of the horn sound matched



Figure 2: An example camera image pair used in the experiment. The left image is the left eye view, the right image is the right eye view. Both images were rendered stereoscopically. There were three such image pairs; one pair for each scene.



Figure 3: A screenshot of the view from within the Oculus Rift DK2. The UI was displayed in all trials as a reminder of the control scheme. It only disappeared from view during each trial interval, after the participant had input their response and before they had begun the next trial at which point it reappeared.

the position of the target car, they confirmed their response using the game controller and moved on to the next trial. Each participant completed 30 such trials, presented in randomised order.

6 RESULTS

All data were processed using the R Language and Environment for Statistical Processing with plots produced using the R graphical statistics package ggplot². Figure 4 shows the results of a regression model, with output of the incongruent function plotted against the perceived distance. The raw data for each set distance is also plotted on a continuous scale. Data is averaged over all 40 participants, with trial repetitions included, governing 20 trials on average. A statistically significant effect was observed, with the model explaining over 91% of the variance ($R^2 = 0.91$, $p < 0.01$). Figure 5 is the residuals plot demonstrating homogeneity of variance.

Table 1 shows the error margins for each distance in meters over all 3 scenes. Distances that resulted in outliers were removed, reducing the number of distance targets in scene 2. Each scene was taken from a different angular perspective, with cars of different shapes, sizes, and color. The response values follow a linear rise, with variations most notable in Scene 2. During debrief, some participants divulged different strategies for matching the horn sound to the target car. These strategies consisted of ‘always picking the front of the car’ to ‘matching the nearest part visible’ to ‘matching the horn to the perceived center of the car’.

²R Project: <https://www.r-project.org/>

Distance	Mean Response	Error	Scene
5.00	0.34	4.66	Scene 1, 2, & 3
7.18	0.50	6.67	Scene 1
7.31	0.283	7.03	Scene 3
7.76	0.45	7.31	Scene 2
9.47	0.68	8.79	Scene 1
9.71	1.00	8.71	Scene 3
10.45	0.61	10.04	Scene 2
11.80	1.06	10.74	Scene 1
12.15	0.90	11.25	Scene 3
13.59	1.16	12.43	Scene 2
14.16	1.11	13.05	Scene 1
14.61	1.34	13.27	Scene 3
16.52	1.572	14.95	Scene 1
17.08	1.62	15.46	Scene 3

Table 1: Table of error margins for mean responses over each distance in the trial set.

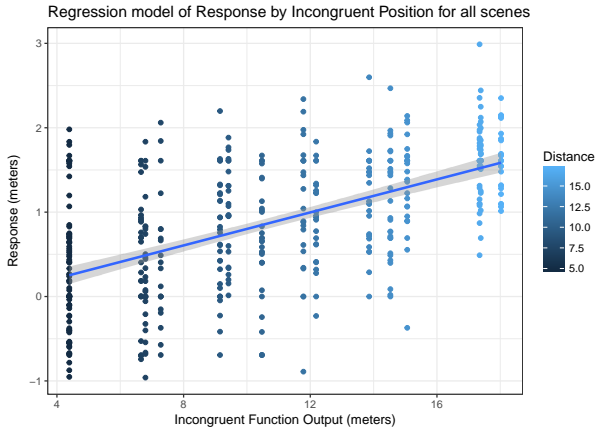


Figure 4: Log-Linear plot of incongruent function output against response values from participants. Actual distances are shaded on a continuous scale. The difference in the axes represents the compression rate of participants over all scenes, demonstrating the need for calibration of the incongruent function.

7 DISCUSSION

The aim of this study was to investigate whether incongruence would be effective in photorealistic environments that provide distance cues to the observer. Finnegan et al. hypothesized that audiovisual distance perception could be manipulated by the application of an incongruence function that would result in the auditory and visual stimuli of an audiovisual target being rendered from different Z locations [9]. While that study was limited to a 2AFC paradigm, our current study provides further evidence in the form of a perceptual matching task. The results from the log-linear regression model on response by incongruence function show a strong correlation of computed versus measured distance perception, but the misalignment of the axes in Figure 4 shows the need for a calibration mechanism. Although participants perceived depth in the scene, they still perceived severe compression. As the incongruence method is based upon fitted curves of perceptual studies, the α and k parameters are decided based on past data. This implies that, as different contexts seem to produce varying compression results [20, 24, 28], the incongruence function itself is context sensitive. Therefore, an

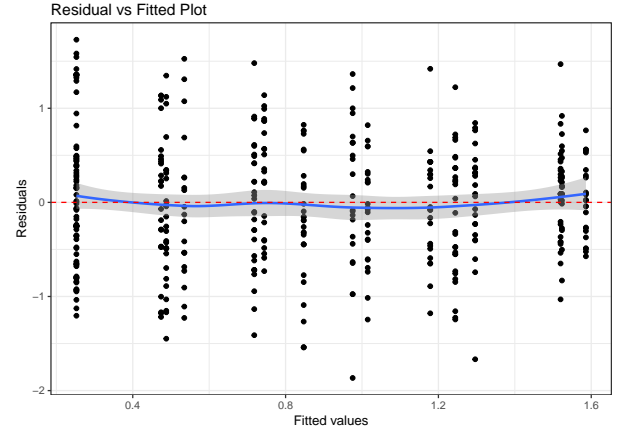


Figure 5: Residual plot of the incongruent function model from Figure 4 demonstrating similar variance across distances.

adaptation, or more appropriate, a ‘training’ period is required to calibrate the incongruence function to the observer’s individual rate of compression. The perceptual matching task used here could be adapted into an appropriate calibration procedure. More research gauging the flexibility of incongruence in other environments will lead to a more complex, general function.

The fact that compression occurred is expected, as numerous past trials attest to this [2, 10]. Severe distance compression has been observed in at least one previous study which also used a head mounted display [29]. It is unclear why such severe compression occurred here, but one reason may be that observers had no reference distance in the visual modality. The only reference given was in the auditory domain consisting of the horn sound positioned 1 meter in front of them. Perhaps participants saw this as a clue to the actual location of the cars, and thus used this as a basis for generating their estimates. Another reason the cars were perceived as being much closer could be due to the binocular disparity kept constant throughout the experiment and across all scenes. Results from Turner et al. suggest that 3D audio may be used to enhance the perception of binocular depth [27]. Combining this evidence with the use of a constant IPD value in the presentation of stimuli in our study, is it viable to suggest that there are subtle crossmodal interaction effects between vision with respect to IPD and binaural audio. However, more research, specifically in the form of a controlled experiment specifically looking at the interaction between IPD and spatial audio position on depth perception in binocular environments, would be needed to confirm this.

As noted, some participants chose different strategies for positioning the car horn. Trials where participants positioned the sound to the front of the car would result in different distance response values compared to participants who chose to just put the sound to the center or rear of the car. Thus asking participants which part of the target object to match to affects the results. Specifying the exact part of the target object to match would lead to reduced error in the calibration phase.

8 CONCLUSION

Previous studies in audiovisual distance perception in VR report compression of distance. Researchers have explained this compression in the context of numerous factors, including response measurement protocol, graphics rendering, direct-to-reverberant ratio of auditory stimuli. We studied distance compression with the goal of *reducing* compression through crossmodal compensation. Our work is based on a theory that distance compression may be compensated through intentionally misaligning the audio and visual stimuli components

of an audiovisual target.

We tested the hypothesis that distance perception can be predicted by an inverted power function whose parameters were based on prior studies in distance compression. We replicated the finding of distance compression measured using a perceptual matching task. Our results showed a log-linear fit for perceived location for a given distance, suggesting that compression is a *systematic* perceptual phenomenon whose parameters may be determined. Determining these parameters has beneficial impact to VR practitioners and designers of VR environments, by recommending a mapping between the acoustic and visual spaces that result in close to real world distance spatial perception.

9 ACKNOWLEDGMENTS

Daniel J. Finnegan's research is funded by CAMERA, the EPSRC Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1. Eamonn O'Neill's research is also partly funded by CAMERA. Michael J. Proulx's research is partly funded by EPSRC grant EP/J017205/1, Design Patterns for Inclusive Collaboration. We thank Soumya Chinnachamy Barathi and Tayfun Essenkaya for their insightful comments.

REFERENCES

- [1] D. Alais and D. Burr. Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3):257–262, 2004. doi: 10.1016/S0960-9822(04)00043-0
- [2] P. W. Anderson and P. Zahorik. Auditory/visual distance estimation: accuracy and variability. *Frontiers in Psychology*, 5:1–11, 2014. doi: 10.3389/fpsyg.2014.01097
- [3] P. W. Battaglia, R. A. Jacobs, and R. N. Aslin. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 20(7):1391–1397, 2003. doi: 10.1364/JOSAA.20.001391
- [4] G. Bruder, P. Lubas, and F. Steinicke. Cognitive Resource Demands of Redirected Walking. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):539–544, apr 2015. doi: 10.1109/TVCG.2015.2391864
- [5] J. S. Chan, C. Maguinness, D. Lisiecka, A. Setti, and F. N. Newell. Evidence for crossmodal interactions across depth on target localisation performance in a spatial array. *Perception*, 41(7):757–773, 2012. doi: 10.1068/p7230
- [6] C. S. Choe, R. B. Welch, R. M. Gilford, and J. F. Juola. The ventriloquist effect: Visual dominance or response bias? *Perception & Psychophysics*, 18(1):55–60, jan 1975. doi: 10.3758/BF03199367
- [7] N. A. Dodgson. Variation and extrema of human interpupillary distance. vol. 5291, pp. 36–46, may 2004. doi: 10.1117/12.529999
- [8] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. doi: 10.1038/415429a
- [9] D. J. Finnegan, E. O'Neill, and M. Proulx. Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 200–212, 2016. doi: 10.1145/2858036.2858065
- [10] M. Gorzel, D. Corrigan, G. Kearney, J. Squires, and F. Boland. Distance Perception in Virtual Audio-Visual Environments. *25th UK Conference of the Audio Engineering Society: Spatial Audio In Today's 3D World*, pp. 1–8, 2012.
- [11] M. Grassi and A. Soranzo. MLP: a MATLAB toolbox for rapid and reliable auditory threshold estimation. *Behavior research methods*, 41(1):20–28, 2009. doi: 10.3758/BRM.41.1.20
- [12] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception*, 7(4):1–18, 2010. doi: 10.1145/1823738.1823744
- [13] B. Hervy, F. Laroche, J.-L. Kerouanton, A. Bernard, C. Courtin, L. D'haene, B. Guillet, and A. Wael. Augmented historical scale model for museums. In *Proceedings of the 2014 Virtual Reality International Conference on - VRIC '14*, pp. 1–4. ACM Press, New York, New York, USA, 2014. doi: 10.1145/2617841.2617843
- [14] G. Kearney, M. Gorzel, F. Boland, and H. Rice. Depth Perception in Interactive Virtual Acoustic Environments Using Higher Order Ambisonic Soundfields. *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics, 2010, May 6-7, Paris, France*, 2010.
- [15] S. A. Kuhl, W. B. Thompson, and S. H. Creem-Regehr. Minification influences spatial judgments in virtual environments. *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization - APGV '06*, 1(212):15, 2006. doi: 10.1145/1140491.1140494
- [16] J. M. Loomis, J. A. Da Silva, N. Fujita, and S. S. Fukushima. Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):906–921, 1992. doi: 10.1037/0096-1523.18.4.906
- [17] J. M. Loomis and J. W. Philbeck. Measuring spatial perception with spatial updating and action. In M. B. Roerta L. Klatzky Brian MacWhinney, ed., *Embodiment, Ego-Space, and Action*, Carnegie Mellon Symposia on Cognition, chap. 1, pp. 1–43. Taylor and Francis, 2008.
- [18] S. Mateeff, J. Hohnsbein, and T. Noack. Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception*, 14(6):721–727, 1985. doi: 10.1068/p140721
- [19] F. O. Matu, M. Thøgersen, B. Galsgaard, M. M. Jensen, and M. Kraus. Stereoscopic augmented reality system for supervised training on minimal invasive surgery robots. In *Proceedings of the 2014 Virtual Reality International Conference on - VRIC '14*, pp. 1–4. ACM Press, New York, New York, USA, 2014. doi: 10.1145/2617841.2620722
- [20] B. J. Mohler, J. L. Campos, M. B. Weyel, and H. H. Bühlhoff. Gait parameters while walking in a head-mounted display virtual environment and the real world. *Proceedings of the 13th Eurographics Symposium on Virtual Environments*, pp. 85–88, 2007.
- [21] T. L. Ooi, B. Wu, and Z. J. He. Distance determined by the angular declination below the horizon. *Nature*, 414(1991):197–200, 2001. doi: 10.1038/35102562
- [22] R. S. Renner, E. Steindecker, M. Müller, B. M. Velichkovsky, R. Stelzer, S. Pannasch, and J. R. Helmer. The Influence of the Stereo Base on Blind and Sighted Reaches in a Virtual Environment. *ACM Transactions on Applied Perception*, 12(2):1–18, mar 2015. doi: 10.1145/2724716
- [23] R. S. Renner, B. M. Velichkovsky, and J. R. Helmer. The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*, 46(2):1–40, 2013. doi: 10.1145/2543581.2543590
- [24] J. K. Stefanucci, K. T. Gagnon, C. L. Tompkins, and K. E. Bullock. Plunging into the pool of death: Imagining a dangerous outcome influences distance perception. *Perception*, 41:1–11, 2012.
- [25] B. E. Stein and T. R. Stanford. Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4):255–266, 2008. doi: 10.1038/nrn2331
- [26] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics*, 16(1):17–27, 2010. doi: 10.1109/TVCG.2009.62
- [27] A. Turner, J. Berry, and N. Holliman. Can the perception of depth in stereoscopic images be influenced by 3D sound? *Displays*, 7863(February), 2011. doi: 10.1117/12.871960
- [28] D. Waller and A. R. Richardson. Correcting distance estimates by interacting with immersive virtual environments: effects of task and available sensory information. *Journal of Experimental Psychology: Applied*, 14(1):61–72, 2008. doi: 10.1037/1076-898X.14.1.61
- [29] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr. Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 17(1):91–101, 2008. doi: 10.1162/pres.17.1.91
- [30] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846, 2002. doi: 10.1121/1.1458027
- [31] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory Distance Perception in Humans: A Summary of Past and Present Research. *Acta Acustica united with Acustica*, 91(3):409–420, 2005.